

Sprache verstehen - auch in explosiven Mischungen



Dr. Jan Rennies-Hochmuth

Dr. Jan Rennies-Hochmuth studierte von 2003 bis 2008 Engineering Physics in Oldenburg (DE), Lyngby (DK) und München (DE). Nach seinem M.Sc. Abschluss schloss er sich der neu gegründeten Fraunhofer-Projektgruppe Hör-, Sprach- und Audiotechnologie in Oldenburg an. Dort promovierte er 2013 zum Thema binauraler Sprach- und Lautheitswahrnehmung und leitet seit 2011 die Gruppe „Persönliche Hörsysteme“ mit Forschungsschwerpunkten in Sprachwahrnehmung, Sprachverbesserung und Psychoakustik.

Kurzfassung des Vortrags No. 42

Sprache verstehen - auch in explosiven Mischungen Immer mehr Kino kommt in die deutschen Wohnzimmer, mehr Abenteuer, mehr Ambiente, mehr Dynamik... in Abhängigkeit der eigenen Hörfähigkeiten und Vorlieben aber auch mehr Unzufriedenheit und Mangel an Sprachverstehen. Nicht jeder liebt laute Explosionen oder reißende Effekte – Hörvorlieben im heimischen Wohnzimmer sind sehr individuell. Die Fernsehsender insbesondere die öffentlich-rechtlichen Kanäle stehen im Sinne der Barrierefreiheit in der Pflicht. Die derzeit präferierte Lösungsstrategie folgt der Idee eine zweite, leichter verständliche Tonspur anzubieten. Diese soll möglichst kurz vor der Ausstrahlung automatisiert erstellt werden, ohne den bestehenden Workflow in der Produktion zu beeinflussen. Den Mix am Ende der Übertragungskette anzupassen ist allerdings nicht ohne weiteres möglich. Eine Adaption im Sinne der Sprachverständlichkeit macht es nötig Sprache und Hintergrund getrennt anzupassen. Die Signale liegen zu diesem Zeitpunkt aber nicht mehr getrennt vor. Hier kommen neuronale Netze ins Spiel, die in der Signalverarbeitung vielseitig eingesetzt werden: Neuronale Netze werden trainiert, um im Mischsignal Sprachanteile zu detektieren, die Verständlichkeit über die Zeit im Signal zu beurteilen, das Signal in Sprache und Nichtsprache aufzutrennen, und einen leichter verständlichen Remix zu erstellen - der Beitrag stellt ein Konzept vor, mit dem sich eine zweite automatisch erstellte Version eines Mixes für Menschen mit Sprachbarrieren ohne Auswirkung auf den Produktionsprozess realisieren lässt. Der Grad der Signaladaption hängt dabei von der momentan geschätzten Höranstrengung (Listening Effort) relativ zur angestrebten bzw. individuell gewünschten Höranstrengung ab. Hierfür wird eine single-ended (d. h. referenzfreie) Methode verwendet, um die Momentan-Höranstrengung zu bewerten. Überschreitet die Höranstrengung einen definierten Wert, so wird das Signal entsprechend angepasst und die Höranstrengung verringert. Die automatische Erstellung der Alternativen Tonspur kann dabei individuell konfiguriert werden und kann somit zur Verbesserung der Sprachverständlichkeit in Rundfunk und Fernsehen beitragen.

Der Vortrag gibt einen Überblick über die Technologie und veranschaulicht die Ergebnisse mit Hörbeispielen.

[PDF anzeigen](#)